



M1-05: Stratification

Part of the "Basics of Data Science with Python" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m1-05/>

More on Confounders and Observational Studies

Lighters and Lung Cancer-- Suppose an observational study showed that people who carry lighters have a higher rate of lung cancer. There are 2 possibilities:

Puzzle #1: Below are either possible confounders, causal links, or neither. Create your own diagram to decide which is which!

a) Genetics: Some people have certain genetics that make them more likely to get lung cancer than others.

b) Smoking Cigarettes: Smoking cigarettes is a known cause of lung cancer. People who smoke cigarettes are also more likely to carry lighters.

c) Lighter fluid: Perhaps inhaling the fumes from lighters can cause lung damage leading to cancer.

***Now how can we determine whether it's the lighters (the treatment) or the smoking (confounder) that is the true cause of lung cancer (the response)?

Stratification

Statisticians adjust for these confounding variables by breaking the control and treatment groups into smaller more homogeneous subgroups, where the confounding factor is the same. This is called **stratification**. Stratification plays a similar role in observational studies as blocking does in randomized experiments and stratification helps us deal with confounders.

Non Smokers	Compare lung cancer rates of those who carry lighters to those who do not.
Moderate Smokers	Compare lung cancer rates of those who carry lighters to those who do not.
Heavy Smokers	Compare lung cancer rates of those who carry lighters to those who do not.



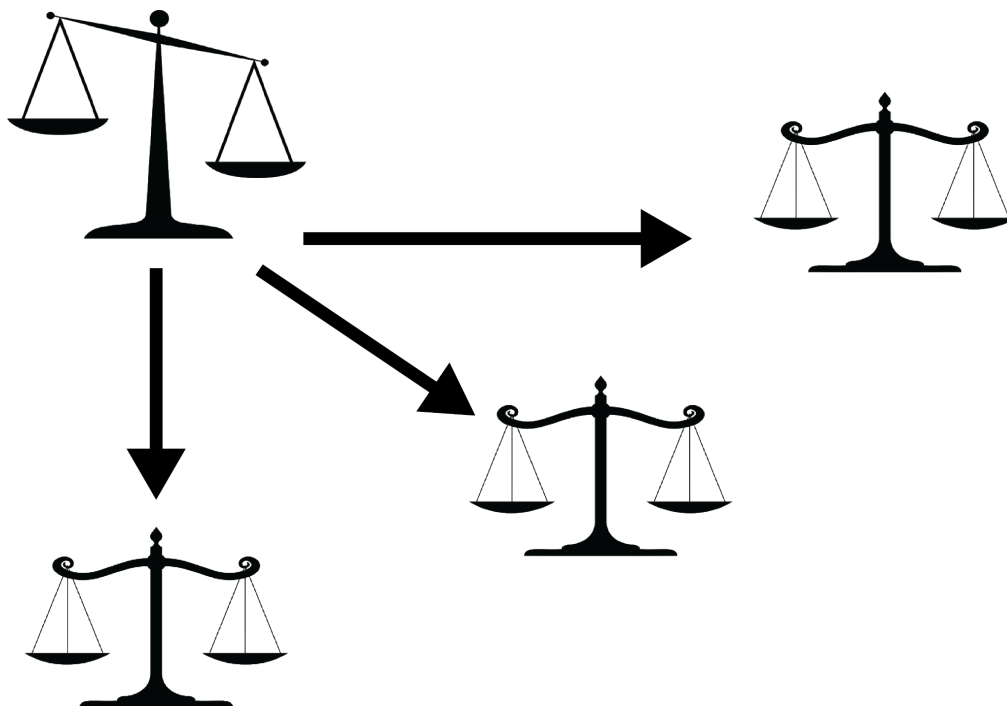
M1-05: Stratification

Part of the "Basics of Data Science with Python" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m1-05/>

In this case, the researchers saw no difference in lung cancer rates between those who carried lighters and those who did not within each group. The heavy smoker group had the highest lung cancer rates, but the rates between those who carry lighters and those who did not are the same in that group (and the other groups).

Hence, smoking is the true cause of the higher rates of lung cancer, not the lighters. We can see this visually below!



With observational studies, you need a much bigger sample size than you do with randomized experiments because each time you stratify for a possible confounder the comparison groups get smaller and smaller.

*How would you stratify to account for a second confounding variable?